Information Operations: Creating New Frontiers for Data Fusion and Mining Technology

Ed Waltz
ERIM International
P.O. Box 134008
Ann Arbor, MI 48113-40008
waltz@erim-int.com 734-994-1200 ext. 2618

The recent developments in the definition of doctrine for information operations and the implementation of systems to enable the conduct of those operations pose new challenges to the application of data fusion and data mining technologies. This paper describes the basic reasoning processes underlying both data fusion and data mining, and discusses the critical role of these processes to enable the development of increasingly complex information operations (IO). The development of IO, this paper asserts, has opened three challenging frontiers of development that must be addressed by data fusion and mining developers. These frontiers include: 1) The need to apply the processes beyond the physical domain, to encompass the symbolic and cognitive domains of reality, 2) The need to extend the processes to deal with subjective and qualitative data, and 3) The need to develop collaborative fusion and mining processes that collaborate with human teams to solve complex problems.

The Elements of Information Operations

The U.S. Joint Vision 2010 statement by the Joint Chiefs of Staff envisions the future capabilities required to maintain military superiority in the 21st century and articulates the role of IO.[1] The basis of the warfighting vision is the achievement of massed effects on an adversary – achieved by dominant maneuver of forces, precision engagement of targets, focused logistics to sustain operations and full-dimensional protection of the forces. Each of these operational concepts is enabled by a state of *information superiority*. Superiority in the information domain is achieved by the conduct of information operations, making this class of operations the critical common denominator to the vision (Figure 1).

The general taxonomy of IO, illustrated in the figure, includes three operational components:

- Dominance – The acquisition of data, organization of information and understanding of the battlespace to maintain dominance in knowledge of own and opponent's forces.
- Defense - The protection of the knowledge obtained by dominance and the assurance that it remains available for all operations.
- Offense - The attack of the opponent's knowledge of the battlespace by denial, disruption, deception, exploitation and destruction of the knowledge and assets used to produce that knowledge.

Form SF298 Citation Data

Report Date ("DD MON YYYY") 00001999	Report Type N/A	Dates Covered (from... to) ("DD MON YYYY")

Title and Subtitle Information Operations: Creating New Frontiers for Data Fusion and Mining Technology	**Contract or Grant Number**
	Program Element Number
Authors Waltz, Ed	**Project Number**
	Task Number
	Work Unit Number
Performing Organization Name(s) and Address(es) ERIM International P.O. Box 134008 Ann Arbor, MI 48113-40008	**Performing Organization Number(s)**
Sponsoring/Monitoring Agency Name(s) and Address(es)	**Monitoring Agency Acronym**
	Monitoring Agency Report Number(s)
Distribution/Availability Statement Approved for public release, distribution unlimited	
Supplementary Notes	
Abstract	
Subject Terms	
Document Classification unclassified	**Classification of SF298** unclassified
Classification of Abstract unclassified	**Limitation of Abstract** unlimited
Number of Pages 13	

The defensive and offensive focus on control of information, while the dominance component focuses on acquisition and understanding – the traditional role of intelligence.

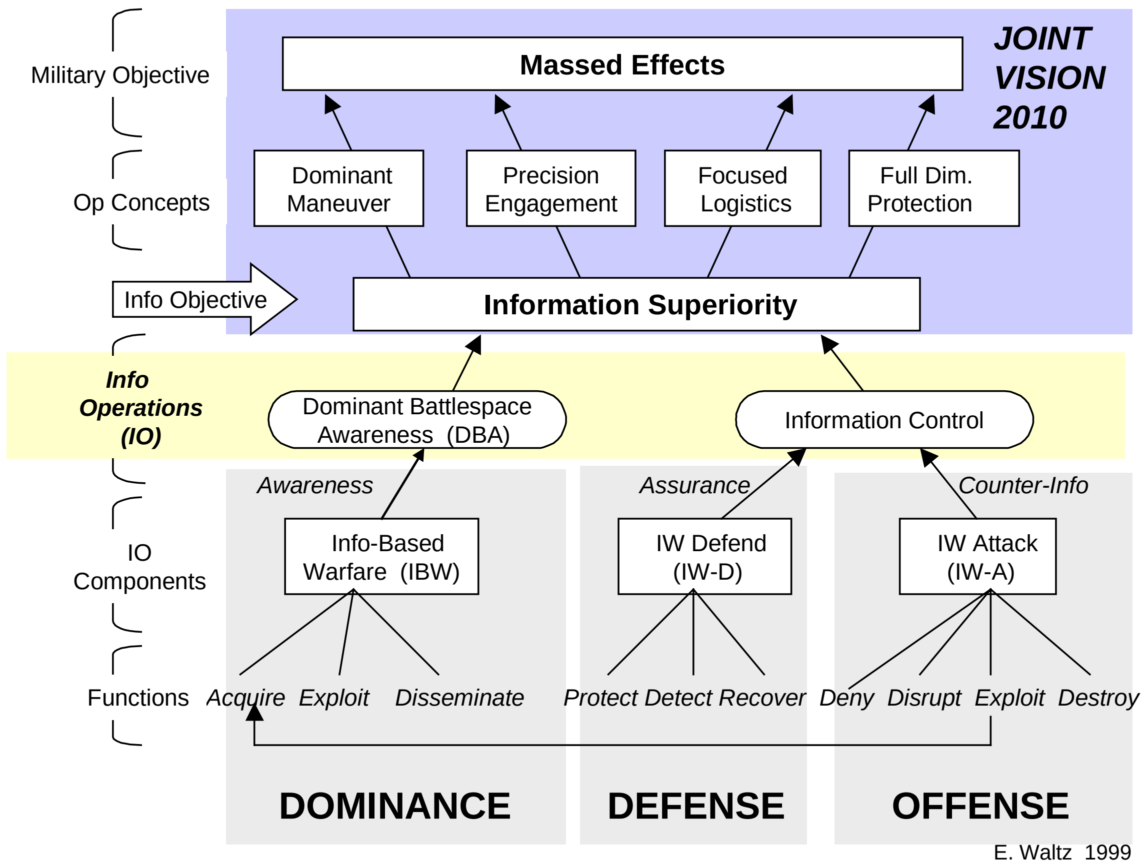

Figure 1- Taxonomy of Information Operations (IO) components and contribution to Information Superiority

Data Fusion and Data Mining in Information Operations

Each of the three elements of IO require support from automated reasoning processes, also referred to as "cognitive services." These reasoning processes include the functions that transform data to information, and information to knowledge. These transformation processes emulate and complement the human cognitive processes, automating three fundamental reasoning or "knowledge creating" activities (Table 1):

- Discovery - These functions discover, or learn unique patterns or characteristics in data. The functions may be supervised (i.e. "training" – the search for common patterns in sets of data known to represent a specific target category) or unsupervised (i.e. the search for patterns in data not necessarily known to contain any targets). Commercial data mining tools perform many of these discovery functions and provide means for semi-automated induction, in which discovered patterns are presented to human analysts. The analyst determines the *meaning* of the pattern- validating it as a discovered signature of a target, for example, or recognizing it as an unusable artifact of the data.

- Detection – The basis of detection is deduction - correlation of a previously known pattern (also referred to as a matched filter template or a signature) to a similar pattern in data to deduce the presence of a target. Data fusion is primarily a deductive process, applying known templates to multiple data sources at multiple levels to deduce signals, objects, situations and impacts (JDL levels 0, 1,2,and 3).
- Explanation – The abductive inference concepts introduced by C.S. Pierce focus on "explanation" of observations by reasoning to create and select the best explanation. This process combines both synthesis (creation of plausible hypotheses) and analysis (assessment of the feasibility of hypotheses to choose the best explanation).

Table 1 The Fundamental Reasoning Processes

	Reasoning Process	**Reasoning Mechanisms**	**Contributing Technology *and application emphasis***
Formal	***Discovery***	*Concept Learning* – Clustering and induction of decision trees	Data Mining –Automated induction processes, discovery of knowledge in databases. Discovery of patterns, signatures and correlations within large databases (information warehouses). across many dimensions of data categories.
		Logic Learning – Production of logical rules	
		Inductive Inference – Develops new knowledge by analogous reasoning that establishes relationships	
	Detection	*Apprehension* (Perception and Conception) – abstraction, extraction of features, and assignment of meaning	Data Fusion – Automated deduction processes, detection of known patterns or signatures in real-time or stored data. Perform alignment, filtering, correlation, categorization and abstraction of large volumes of raw (or preprocessed) data.
		Judgment – Construction and adjudication of propositions	
		Deductive Inference – Provides new knowledge by establishing relationships between propositions	
Common-Sense	***Explanation***	*Explanation* – Generalization to develop a fruitful family of explanatory hypotheses that explain observations	Information Understanding- Automated synthesis, ranking and selection of comprehensive sets of plausible explanations of evidence. Analogous and case-based reasoning, with the evaluation of multiple perspectives and collaborative reasoning across diverse views.
		Abductive Inference – Creates new relevant hypotheses to explain evidence with provision to select most plausible one(s)	

These reasoning functions are automated to "augment the human intellect" [2] of the IO commanders and warfighters to comprehend complex information situations, and to plan/respond rapidly. They are applicable to all three components of information operations (Table 2)– not just in the typical intelligence functions performed in the dominance operations.

The table illustrates representative application areas in each of the three components of information operations, especially in the area of information network operations:

- Dominance –The intelligence collection and understanding functions require discovery to locate and identify new threat networks, their behaviors and target signatures. These signatures are then used for continuous detection and tracking of target nets, nodes, traffic and situations in areas of interest.

- Defense – Defensive operations, similar to dominance, require the detection of incipient and imminent threats and intrusions. The methods of data fusion and data mining are being applied to network defensive systems to analyze (mining) and combine-to-detect (fusion) intrusion patterns of complex multiple-coordinated attacks across networks.
- Offense – In support of targeting, weaponeering and damage assessment, reasoning processes in software tools provide support to human planners today. As offensive information operations become increasingly automated and autonomous to meet the rapid response times of information attacks, reasoning functions will be required to allow weapons to seek targets, identify vulnerabilities, and attack based on learned vulnerabilities.

Table 2- Representative Roles of Reasoning Processes within Network Information Operations

	DOMINANCE	***DEFENSE***	***OFFENSE***
Detection *(Deductive Processes)*	• **"Target" detection and tracking (three DF levels, three domains)**	• **Indications and warning** • **Network Intrusion detection**	• **Targeting analysis** • **Weaponeering** • **IO damage assessment**
Discovery *(Inductive Processes)*	• **Signature discovery** • **Behavior analysis** • **Cryptanalysis**	• **Intrusion pattern discovery** • **Intruder tracing and identification**	• **Inquisitive weapons (e.g. "discovery virus")**
Explanation (Abductive Processes)	• **Situation explanation (DF level 2) and impact prediction (DF level 3)**	• **Threat explanation and reporting** •	

Three Frontiers for Data Fusion and Data Mining

The introduction of information superiority as a principal military objective and the achievement of that objective by means of Information Operations introduce two significant implications for the developers of traditional data fusion systems. These are:

1. IO requirements are seeking data fusion solutions that combine data that is neither objective nor quantitative, and that address targets that are not physical. IO systems require fusion and mining processes that reason about abstract targets (e.g. information itself) in domains other than the physical.
2. Data fusion and data mining are complementary processes that must be integrated to provide robust deductive-inductive reasoning support to IO. The defense community that has developed and applied data fusion technology must integrate and apply the

commercial data mining technology that has been largely developed and applied in the commercial world.

The full meaning of these implications can be best described by presenting the three challenging frontiers that now face the data fusion development community. The following sections introduce each of the frontiers of challenge in the context of intelligence operations, although the applications of these capabilities extend across the dominance, defense and offense components of IO. These frontiers must be conquered in order to provide the enabling fusion technology to support the Information Operations envisioned by the creators of Joint Vision 2010.

Frontier 1: Extend Reasoning to All Three Domains of Human Reality

To date, the emphasis of data fusion development has been focused on the immediate military problem of locating and identifying military targets – stationary and moving objects. Using all sources of data, the fusion process has correlated sensed observations and applied the constraints of correlated observations and physical laws to model the behavior of the target objects. This ability to detect, identify and track physical objects, must be extended to detect, track and identify objects in two other domains of interest to information operations.

Consider a three-domain view of human reality that is of keen interest to IO planners and warfighters:

- The ***Physical*** domain includes physical military objects: facilities, lines of communication, vehicles, aircraft, missiles, and personnel. The "orders of battle" that measured Cold War military strength were determined by counting missiles, warheads, tanks and trucks -- all objects of the physical world.
- A more abstract domain, though, is the ***symbolic*** domain -- the realm of information. Words, numbers, graphics, all encode and represent the physical world, storing and transmitting it in electronic formats, radio and TV signals, the Internet, newsprint and other forms. This is the domain that is expanding at unprecedented rates, as global ideas, communications and descriptions of the world are being represented in this domain. The domain, also described as the “infosphere” or “cyberspace” has become the principal means by which humans shape their perception of the world.
- The ***Cognitive*** domain is the realm of human thought. This is the ultimate locus of all information flows. The individual and collective thoughts of government leaders, and populations at large form this realm. Perceptions and decisions -- and the effects on our nation are formed in this cognitive realm. This is the ultimate target of our adversaries: the realm where uncertainties, fears, panic and terror can coerce and influence our behavior.

These three distinct domains (Figure 2) are each characterized by intelligence objects (targets) and roles in human activity. The domains provide a functional model for understanding and organizing intelligence operations (exploitation of information), as well as offensive and defensive information operations.[3] While the methods of representing (modeling) and analyzing the physical realm and the symbolic realm are well known, the development of methods to represent the cognitive domain remains a significant challenge for next-generation intelligence

architectures. The symbolic and cognitive domains are the new challenges for “target” modeling and explanation.

DOMAIN	Disciplines of Study	Objects of Analysis	ROLE	Direct Intelligence
Cognitive *(Synapseware)* the Human Mind and heart	•**Epistemology** •**Psychology** *Issues of existence and consciousness*	•**Human will** •**Ideas, thoughts, knowledge**	**Will** **Decide** **Perceive**	
Information *(Software)* abstract symbolism and electronics	•**Logic** •**Information theory** *Issues of symbolic representation, manipulation (math and language) and knowledge*	•**Symbols, uncertainty,**	**Compute** **Transfer** **Symbolize**	•**SIGINT** •**NETINT** •**HUMINT**
Physical *(Hardware)* matter	•**Physics** *Issues of matter and energy*	•**Mass, energy**	**Sense**	•**IMINT** •**MASINT** •**HUMINT**

Figure 2- Three Domains of Human Reality from an Intelligence Perspective

Representative examples of the objects (or targets), groups, and states that must be modeled and explained in all three domains are summarized in Figure 3 to illustrate the breadth of coverage of the model to explain human activity.

	DOMAIN	TARGET OBJECT	TARGET GROUP	TARGET STATE
EMPHASIS OF NATIONAL INTELLIGENCE	Cognitive	• People (thoughts) • Belief • Understanding	• Ideological group • Political group • Nation-state	• Belief state • Uncertainty • Strength of will • Intent
	Symbolic	• People (statements, conversations) • Knowledge • Information • Data	• National discourse • Statement group • Conversation group	• Plan • Database • Process
EMPHASIS OF MILITARYINTEL	Physical	• Facilities • Vehicles • People	• Nation-state • Unit • Base	• Kinematic state • Location

Figure 3- Representative Targets, Groups and States in Three Domains

Intelligence, or knowledge, is the product resulting from collection, processing, integration, analysis, evaluation and interpretation of data to explain and even predict events or states of human reality. Current intelligence gathering has organized its efforts along the lines of sources and collection: Imagery Intelligence (IMINT) has focused on observing physical objects, Signals intelligence (SIGINT) has collected symbolic information transmitted across the electromagnetic spectrum, and human intelligence (HUMINT) has collected physical and symbolic data, while also attempting to understand the thoughts of our adversaries. Measurement and Signatures intelligence (MASINT) and other specific technical means have also focused on the physical and symbolic domains.

Future information superiority must provide an infrastructure and information architecture to develop a comprehensive understanding of our adversaries – in all domains. We must be able to know and forecast:

- ***What they do in the physical domain,***
- ***what they say, plan and store in the symbolic domain,***
- ***and what they perceive, conceive and decide in the cognitive domain***

Vice Admiral Arthur Cebrowski, speaking at an information warfare conference[4] has noted the importance of adding the symbolic (information) and cognitive dimensions to our current capability to understand an adversary's physical weapon system behavior:

1. The True focus of Information Superiority is Human cognitive behavior: "We're not looking for the location of a tank, but for the factors that are prime determinants of outcome: cohesiveness, resolve, morale."
2. The focus of DoD RDT&E is NOT on cognitive behavior: "That warfare is dominated by human factors is not appreciated by the DoD's science and technology community. It is steeped in physics and arms of engineering. This [human cognitive dominance] was recognized by Clausewitz, but not by the Pentagon."
3. To achieve Information Superiority, we must invest in and understand this area:"[We should] refocus Science & Technology on the behavioral component of warfare."

We can apply this 3 domain concept to develop a data fusion architecture which is complaint with the U.S. DoD Joint Directors of Laboratories Data Fusion Subpanel model[5] by simply distinguishing the models of the three domains of knowledge being created by the process (Figure 4). The architecture distinguishes:

- Fusion- the fusion processes at three JDL levels (L1 fusion of objects, L2 fusion of groups, and L3 fusion of impacts). In this model the JDL process flows from detection of objects, to detection of groups and situations by relationships and behavior, to detection of impacts by influences in context.
- Models - The dynamic models of reality being maintained at three JDL levels (M1 models of objects, M2 models of groups and M3 models of impacts, as illustrated in Figure 3).
- Inference between Domains- The fusion over three domains of reality in which (at each JDL level) there exists upward –physical to cognitive- and downward –cognitive to

physical- inference linkages. These linkages allow reconciliation, for example, between the observed physical state of a military force, the command information that that it is exchanging that measures its symbolic state, and the perception (cognitive state) of it's military commanders of it's situation.

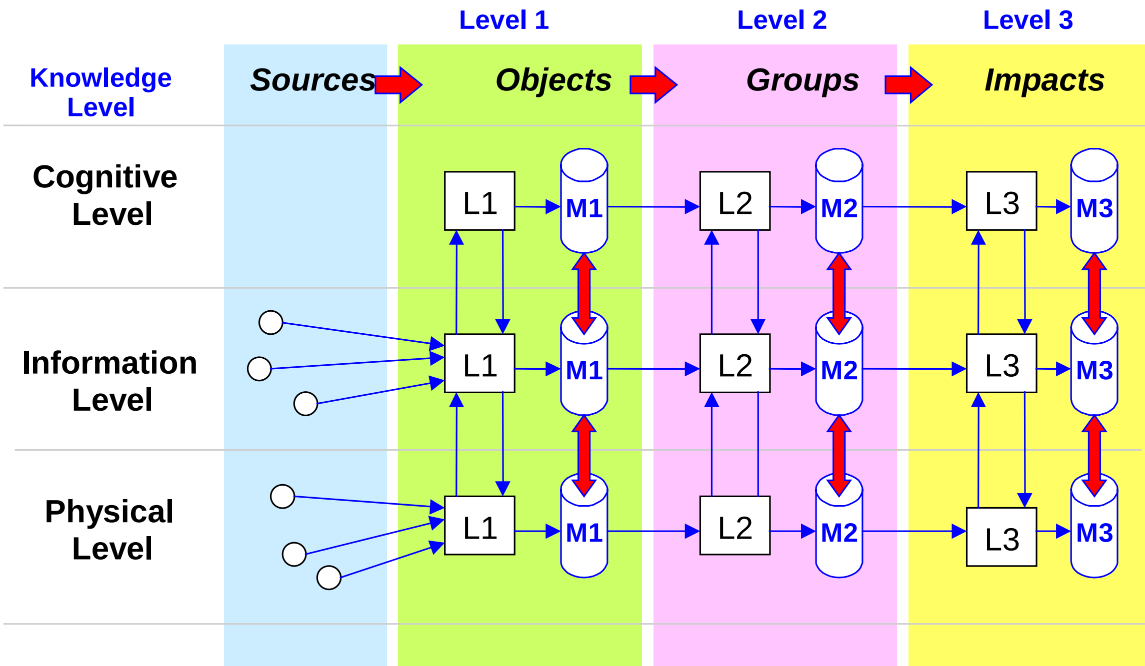

Figure 4 – A Three-Domain Data Fusion Architecture

Frontier 2: Extend Data Fusion to Accommodate Subjective, Qualitative Data

The principal emphasis of DoD data fusion applications has focused on correlating and combining objective, quantitative data. Sensor data is quantified into numerical target reports, and the fusion process applied quantitative methods to align, correlate, combine and model the state of targets in a quantitative sense. The commercial knowledge management industry and the national intelligence community have focused R&D on the development of tools to correlate and combine qualitative text data – using words, phrases and concepts as the basis to search, correlate, combine and abstract from the corpus of electronic texts. Information Operations, especially in the symbolic and cognitive domains, require the ability to combine and model structured and unstructured text data (across multiple languages). The DARPA Dynamic Multiuser Information Fusion (DMIF) program developed message parsing capabilities to convert and extract quantitative data sets (target vectors) from structured tactical report formats. Commercial tools developed by Excalibur[6] and Autonomy[7] are pioneering the manipulation of unstructured text, audio and video data to perform fusion functions that approach those defined in the JDL fusion model, including level 1 fusion of words, topics and concepts.

Data fusion developers must consider approaches to perform fusion of both qualitative and quantitative data to develop understandings of situations in which both categories of data are available. Combined fusion processes (Figure 5) will allow sensed data (quantitative) and source data (most often qualitative) to be combined to provide a complete understanding of complex problems.

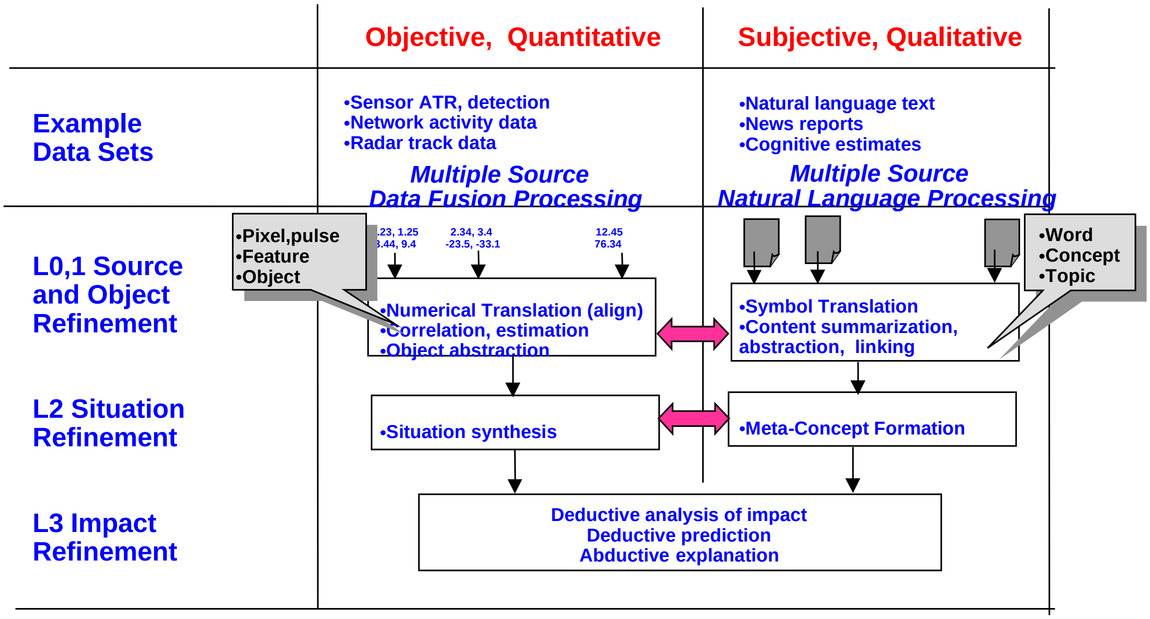

Figure 5- Combined Quantitative-Qualitative Data Fusion

Consider, for example, the problem of assessing the national stability situation of a third world nation in which the following data sets must be combined to derive a model of political and economic stability. Both categories of data types provide valuables sources at the physical and symbolic levels (Table 3) for modeling and describing the observed nation at three levels.

Table 3- Data Sets Available to Monitor National Stability

	Qualitative Data	Quantitative Data
Cognitive	No direct sources	No direct sources
Symbolic	• Government activity levels • Military status	• News and wire reports • Local Radio-TV-Newspapers
Physical	• Imagery-derived industrial production, export data • Imagery-derived	• Government economic reports

While the table list no direct sources of cognitive data, the cognitive states (of national leaders, political and industrial groups, and the population at large) may be able to be inferred from the other two levels.

Knowledgeable subject area analysts currently tackle such problems, but the increasing deluge of global qualitative and quantitative data makes it difficult for those analysts to consider and assess all available data. Combined qualitative-quantitative data fusion and mining technologies will allow all available data to be related and analyzed to bring to the human analysts the most relevant 3-domain model implications, and to allow the analysts to drill-down to the most significant supporting data.

Frontier 3: Integrate Human-Machine Collaborative Reasoning

The third frontier is the need to fully integrate the automated reasoning processes with collaborative human teams to solve complex problems of understanding. The data fusion process, in particular, must be viewed as a process that extends across humans and machines, in collaboration, to perform collective reasoning.

In a recent independent assessment of national intelligence capabilities, the importance of "collaboration" across information flow ("stovepipes") was highlighted in recommendations provided by Adm. David Jeremiah to the intelligence community:

> Look at establishing effective mechanisms to guarantee stronger integration of the analysis and greater collaboration and coordination of intelligence agencies and disciplines. So that instead of looking up at each of these stovepipes, we look at the product and the interaction between the stovepipes. [8]

The challenge for data fusion and mining designers is to allow these processes to effectively enable collaborative virtual teams (working virtually across time and space) to comprehend complex situations from large volumes of data. Fusion and mining capabilities must support Information Operations conducted by virtual enterprise organizations – warfighters operating across the globe as virtual teams, though not necessarily located in the same place, nor communicating at the same time.

One architectural model for such a virtual enterprise for collaborative problem solving that requires machine and human detection, discovery and explanation is depicted in Figure 6. The layered model is related to the three architectural "views" identified by the DoD: 1) Operation View describes the layers of the human work flows and business processes of virtual teams, 2) System View describes the layers of the information system environment that provide computing and interconnection, and 3) Technical View identifies the component standards, protocols and technologies that enable the physical and software implementation.

The layered model includes a hierarchy of metrics that measure utility, effectiveness and performance, as depicted in Figure 6:

- Intelligence Supply Chain- the flow of data-to-knowledge and the flow of value are described as the highest, and most abstract, level of the model. The metric for valuing each function and workflow identification is Return on Information (ROI).
- Virtual Intelligence Enterprise Organization- The collection of all virtual workgroups makes up the organization, including human intellectual capital and organizational knowledge.
- Virtual Workgroups – Individual workgroups (virtual intelligence units, or cells) that dynamically form and dissolve as needs arise are the fundamental units of knowledge creation, and are measured by measures of effectiveness (MOEs).

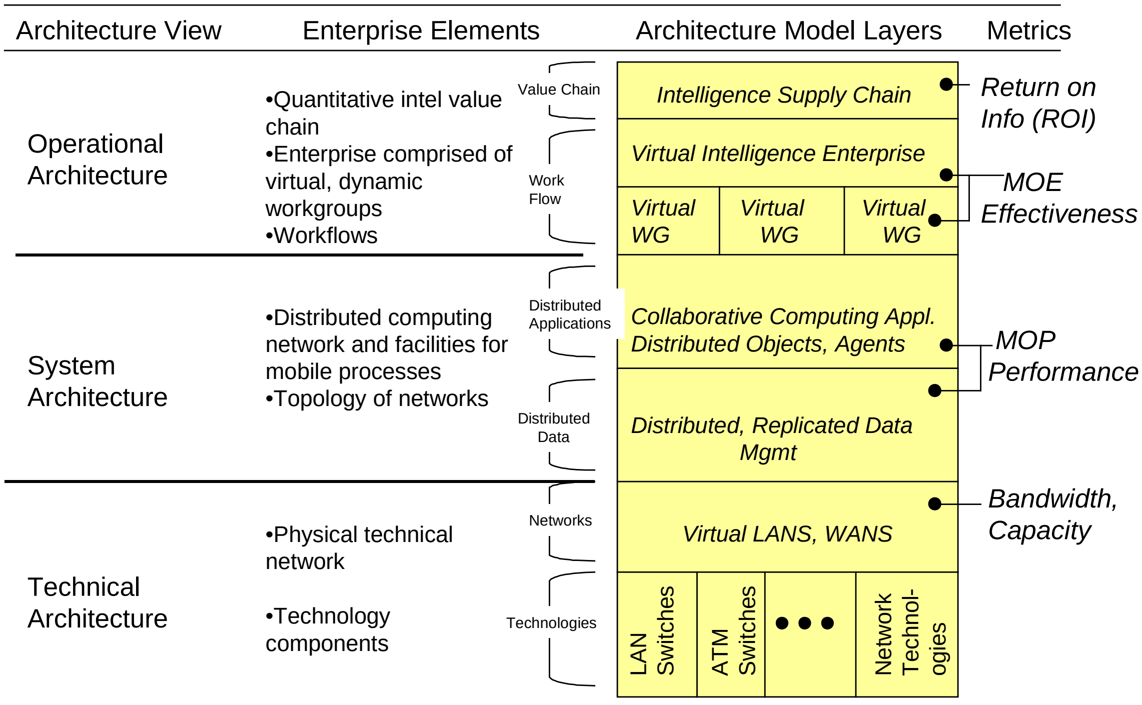

Figure 6: Collaborative Enterprise Architecture Model Layers

- Collaborative Computing Architecture – The collaborative process is performed across a distributed computing architecture, which is measured by performance-level metrics (MOPs).
- Distributed, replicated data management – A layer of distributed data management, with replication capability is the foundation to support distributed computing.
- Virtual Networks and protocols- the technical physical (hardware), and abstract (or software) elements are described by protocols, standards for this layer.
- Technologies – Finally, physical component technologies form the final layer.

A virtual enterprise for Information Operations is depicted in Figure 6 in an N-tier architecture in which collaborating clients interact, via normal web facilities, with a distributed processing layer that accesses the multiple streams and stores of data collected.

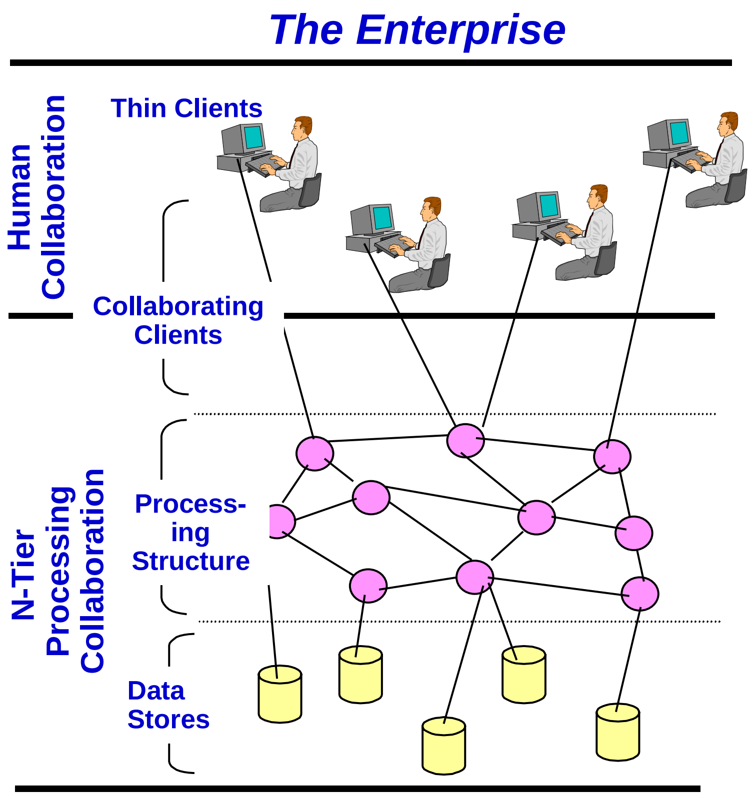

Figure 6- A Typical 3-Tier Collaborative Enterprise Implementation

Summary

Information Operations require significant reasoning support to discover, detect and explain the increasing complex activities of human and machine events that make up human competition and conflict. Reasoning capabilities support the dominance (intelligence), defensive and offensive aspects of information operations. These capabilities, enabled by data fusion and mining technologies, must explain and enable information operations across all three domains of human reality: the physical, symbolic, and cognitive domains. In order to meet the challenge to support these operations, the data fusion community must: 1) Apply the fusion and mining processes beyond the physical domain, to encompass the symbolic and cognitive domains of reality, 2) Extend the processes to deal with subjective and qualitative data, and 3) Develop collaborative fusion and mining processes that enable collaboration collaborate machines and human teams to understand e complex situations and solve complex problems.

Endnotes

[1] Joint Vision 2010, US DoD Joint Chiefs of Staff, U.S. Government Printing Office, 1977.

[2] Engelbart, Douglas C. "Augmenting the Human Intellect: A Conceptual Framework" Summary report AFOSR-3223, Stanford Research Institute, Menlo Park, CA, October 1962. Page 1: "By 'augmenting the human intellect we mean increasing the capability of a man to approach a complex problem situation, to gain comprehension to suit his particular needs, and to derive solutions to problems."

[3] Waltz, Edward, "Information Warfare: Principles and Operations", Artech House, 1998, chapter 5.

[4] Comments were made at the Technology Training Corporation Information Warfare Conference, 16 October 1998, in Washington DC.

[5] Most recent version is described in, Functional *Description of the Data Fusion Process*, Joint Directors of DoD Laboratories, November 1991. At the time of this writing, the model is in revision and refinement by the JDL Sub-Panel on Data Fusion.

[6] Online site excalib.com

[7] Online site autonomy.com

[8] CIA Press Release of News Conference, *2 June 1998,* Adm. David Jeremiah Independent Evaluation of the actions taken by the Intelligence Community leading up to the Indian nuclear test.